Error reduction in density estimation under shape restrictions

Running head: ERROR REDUCTION IN DENSITY ESTIMATION

Tomasz RYCHLIK

Polish Academy of Sciences

Key words and phrases: Nonparametric density estimation, monotone density, symmetric unimodal density, weighted integrated squared error, error reduction, projection.

AMS 1991 subject classifications: 62G05, 62G07.

ABSTRACT

For the problems of nonparametric estimation of nonincreasing and symmetric unimodal density functions with bounded supports we determine the projections of estimates onto the convex families of possible parent densities with respect to the weighted integrated squared error. We also describe the method of approximating the analogous projections onto the respective density classes satisfying some general moment conditions. The method of projections reduces the estimation errors for all possible values of observations of a given finite sample size in a uniformly optimal way and provides estimates sharing the properties of the parent densities.

RÉSUMÉ

L'auteur s'intéresse au problème de l'estimation non paramétrique de fonctions de densité non croissantes ou unimodales et symétriques à support fini. Il détermine la projection d'estimateurs non paramétriques sur des familles convexes de densités de lois par rapport à l'erreur quadratique pondérée intégrée. Il décrit en outre une méthode d'approximation de projections analogues sur les classes de fonctions de densité dont les moments satisfont à certaines conditions générales. Cette technique de projection réduit de façon uniforme et optimale les erreurs d'estimation pour toutes les valeurs possibles des observations d'un échantillon fini donné, en plus de produire des estimations qui partagent certaines des caractéristiques de la famille de lois choisie.

1 INTRODUCTION

Suppose that we wish to estimate the density function of a lifetime distribution knowing that the lifetime is bounded by a given constant and that its density is nonincreasing. If prior knowledge is insufficient to select a parametric model, the density can be estimated nonparametrically. Standard techniques (e. g., the kernel method) often assign some probability to the complement of the support (especially to small negative values). In addition, they do not allow for control of the shape of the estimate.

In the paper, we determine the projection of a given estimate onto the class of possible densities with respect to the weighted integrated squared error ($WISE$, for short). Moreover, we present the solution of the respective problem for the class of unimodal symmetric densities and describe procedures approximating the projections for the analogous problems under several moment constraints. The projection improves the estimator accuracy for all possible values of observations and a given sample size and this is the uniformly optimal correction of the original estimate.

The problem of nonparametric estimating of nonincreasing density was first considered by Grenander (1956) who proposed a maximum likelihood estimate. This is the derivative of the smallest concave majorant of the empirical distribution function, and takes the form of a nonincreasing stepwise function with jumps at some observation points, vanishing at the maximal one. The asymptotic representation of the MLE and its accuracy were established in Prakasa Rao (1969), and in Groeneboom and Pyke (1983) and Groeneboom (1985), respectively. A nonasymptotic analysis of the L_1-risk of the estimate was carried out in Birgé (1987a, 1989). Recently, Fougères (1997) proposed another approach to the problem, modifying a pilot kernel estimate by a monotone rearrangement. She proved that the modification is almost uniformly strongly consistent, and it improves the pilot in L_p-norms, $p \geq 1$, almost surely, and determined its rate of convergence and asymptotic distribution. The Grenander and Fougères estimates have natural generalizations for unimodal densities with a known mode.

Birgé (1997) introduced a modification of the former one for problems with an unknown mode that has the same asymptotic L_1-risk as the original estimate based on the knowledge of the mode. Barlow and van Zwet (1970, 1971), studying more general problems (e. g. estimating monotone failure rates), proposed a monotone histogram-type estimate with bins wider than spacings which tends to the parent density faster than the Grenander estimate. Birgé (1987b) constructed a histogram with geometrically increasing bin widths proving that the estimate is minimax optimal for nonincreasing densities under an appropriate choice of the rate of the width increase. Different approaches to the monotone density estimation problem were discussed in Devroye (1987, Chapter 8).

Kiefer and Wolfowitz (1976) proved that the antiderivatives of the Grenander estimates are asymptotically efficient in estimating concave and convex functions. Deely and Kruse (1968) constructed the Kolmogorov metric projections of the empirical distribution function onto mixtures of distributions (due to Kintchine (1938), every nonincreasing life density is a mixture of scaled uniforms). More general minimum distance estimates for abstract mixture models were studied in Barbe (1998).

Shape restrictions were also treated in regression problems, especially with deterministic designs (see Wright (1982), Friedman and Tibshirani (1984), Mukerjee (1988), Mammen ($1991a, b$)). The methods proposed consist in adapting either data or estimates by projecting onto respective ordered families. The idea of applying spline smoothers satisfying shape constraints were developed by Utreras (1985), Villalobos and Wahba (1987), and Mammen and Thomas-Agnan (1998). For a recent review of nonparametric estimation problems under shape restriction we refer the reader to Delecroix and Thomas-Agnan (1998). A comprehensive study of inference methods in order restricted models can be found in Robertson *et al.* (1988).

A method of projecting improper density estimates onto the family of proper densities in the weighted L_2-norm was proposed by Gajek (1986) and further improved in Gajek (1987). Lenic (1994) studied a similar problem for the densities satisfying a moment condition. Schuster (1987) considered the problem of projecting densities onto symmetric ones. Kałuszka and Lenic (1995) determined the projections onto density functions with respect to the Hellinger distance. The idea of projecting in density estimation can be considered as a special case of the nonparametric minimum distance estimation theory introduced in Vapnik and Stefanyuk (1978) and Stefanyuk (1979), and developed in Vapnik (1982) (see also Devroye (1987, Chapter 6)) where the emphasis was laid on smoothness conditions rather than shape ones. Minimum distance density estimates, especially ones based on the Hellinger norm, proved to be both asymptotically efficient and robust in many parametric models. We only refer here to a pioneering paper of Beran (1977) and a recent review of Basu *et al.* (1997).

Sometimes objections are raised that projections do not improve the rate of convergence of the original estimate. E. g., the rates of convergence of the Grenander estimate and the histogram are identical, and so neither of them is recommended for estimating smooth density functions. The estimates of Barlow and van Zwet (1970, 1971) have the same asymptotic behavior as their unconstrained nonmonotone versions. Kiefer and Wolfowitz (1976) proved that the smallest concave majorant does not improve the asymptotics of the empirical distribution. Barbe (1998) described two types of necessary and sufficient conditions for minimum distance distribution estimators to behave asymptotically in mixture models as the empirical distribution. Similar conclusions for regression models were presented in Mammen ($1991b$) and Mammen and Thomas-Agnan (1998). This phenomenon has a natural explanation. Reasonable unconstrained estimates are consistent and eventually share the properties of the curve being estimated so that the corrections become smaller as the sample size increases.

Nevertheless, we defend projections referring to our problem of estimating a monotone density. Efficient smooth curve estimates based on expansions applied for estimating densities do not preserve the property of being density functions themselves and have to be corrected anyway in order to become meaningful for practical interpretation. It is therefore justifiable to determine first highly efficient unrestricted estimates for models with restrictions and then project them onto the restricted families. This provides estimates with desired properties and improves the accuracy of the original ones for given finite data which is of actual interest in practice. It is worth pointing out that actual corrections valid for every finite sample and every model distribution are derived by projections onto convex sets of Hilbert spaces. For instance, projecting with respect to the Hellinger distance we can consider density functions which are nonnnecassarily square integrable, but the square roots of the densities do not form convex sets and, in consequence, we cannot obtain uniform finite-sample corrections.

In Propositions 1 and 2 of Section 2 we establish the projections of density estimates onto the convex families of nonincreasing and symmetric unimodal densities satisfying some support restrictions. The projections are described up to a single constant that can be determined numerically. For the analogous problems with moment constraints, we show that the respective projections are approximated by sequences of functions that are the projections of combinations of the original estimates with elements of some given finite dimensional subspaces onto specified convex cones (see Propositions 3 and 4). The main tool for the proofs that are presented in Section 3 is an infinite algorithm of Boyle and Dykstra (1986) combining the projections onto a number of convex sets so as to obtain sequences tending to the projection onto the intersection. Adapting the algorithm to a properly chosen partition of conditions on the parent density function we are in a position to guess the final forms of the projections described in Propositions 1 and 2. This is not possible when we additionally assume that the expectations of some generally unspecified functions are given. However, we are able to specify simple algorithms (see Section 4) so that the outcomes of all stages improve the estimation accuracy of the previous ones, and, accordingly, those of the original estimates. The sequences tend to the projections, although they do not belong to the respective classes of possible parent densities. We also specify algorithms for some related problems.

2 RESULTS

In order to estimate a one-dimensional density function f, we carry out an experiment, get some observations X and determine an estimate $\tilde{f} = \tilde{f}(X, \cdot)$. To fix the ideas, we could think of X as of a sequence of a given length of independent identically distributed random variables with the parent density f, but more general models for which f is identifiable (e. g., stationary processes) can be considered. We measure the accuracy of estimation by means of weighted integrated squared error

$$WISE(\tilde{f}, f) = \int [\tilde{f}(X, x) - f(x)]^2 w(x)\, dx, \tag{1}$$

for some nonnegative weight function w, assuming that both f and $\tilde{f}$ for all possible observations are square integrable with the weight, i.e.

$$f, \tilde{f}(X, \cdot) \in \mathcal{H} = L_2(R, w(x)\, dx) = \{h : \int h^2(x) w(x)\, dx < \infty\}.$$

We first assume that the parent density function f satisfies the following

$$f \in \mathcal{F} = \{h \in \mathcal{H} : \quad \begin{array}{l} h \geq 0,\ \int h = 1,\ h = 0 \text{ on } I^c, \\ h \text{ is nonincreasing on } I\}, \end{array} \tag{2}$$

where $I = [0, a]$ is a finite interval and I^c stands for its complement, and $\int_I w$, $\int_I 1/w$ are finite. This implies that for all possible X the integrals $\int_I \tilde{f}$ and $\int_I \tilde{f} w$ exist and are finite. In particular, we do not demand that $\int \tilde{f} = 1$ nor $\tilde{f} \geq 0$. This allows us to use, e. g., efficient kernel estimates with high order kernels or ones based on orthogonal series expansions with a large number of summands. The essential assumptions on w concern its behavior on the support I of the parent density function. The choice of its shape there reflects the importance we put on accuracy of estimating f in some regions with respect to the other ones. For instance, one could be more interested in the higher estimation precision either at the borders (especially at the left border, where the greatest probability mass is concentrated) or in the central part of the domain. The exact form of w in I^c plays no role in our approach and we can always define it there to ensure $\int_{I^c} \tilde{f}^2 w < \infty$.

Our purpose is to determine projection $Q\tilde{f}$ of an arbitrary realization of the estimate $\tilde{f} = \tilde{f}(X, \cdot)$ onto $\mathcal{F}$ with respect to $WISE$, i. e., to find $Q\tilde{f}$ that minimizes the distance (1) among all elements of the convex set (2). This method has two advantages. First, the new estimate is actually a density function that satisfies the shape and support postulates. Second, for every $h \in \mathcal{F}$ (and so for the actual density f in particular)

$$WISE(\tilde{f}, h) \leq WISE(Q\tilde{f}, h). \tag{3}$$

Since this holds for all possible X, by taking the expectations of both the sides of (3) with $h = f$ we conclude that using the projections we decrease mean weighted integrated squared error

$$MWISE(\tilde{f}, f) = \mathrm{E} \int [\tilde{f}(X, x) - f(x)]^2 w(x)\, dx.$$

It can be also verified that in some sense $Q\tilde{f}$ is the uniformly optimal correction of $\tilde{f}$ in $\mathcal{F}$, i. e., for any other modification $Q'\tilde{f} \in \mathcal{F}$ of $\tilde{f}$, depending on data X

through $\tilde{f}$, the following holds

$$\inf_{h\in\mathcal{F}}[WISE(\tilde{f},h) - WISE(Q'\tilde{f},h)] \leq \inf_{h\in\mathcal{F}}[WISE(\tilde{f},h) - WISE(Q\tilde{f},h)].$$

Observe that the idea of reducing the estimation error by projections is not so easily justified in the case of nonconvex classes of model densities. Then (3) does not hold for all possible h and so it may happen that the projection would actually increase the error. It makes some sense, however, in the asymptotic setup.

In order to describe $Q\tilde{f}$, we first introduce some auxiliary notation. Let $g_{|A}$ denote the function coinciding with g on a set A and vanishing elsewhere. Set $g_+ = g_{|\{g>0\}}$. Let W stand for the primary of $w_{|I}$, i.e.

$$W(x) = \int_0^x w(t)\,dt, \qquad x \in I,$$

and let $A = W(a)$. Since W is strictly increasing and absolutely continuous, its inverse $W^{-1} : [0,A] \mapsto I$ exists and has the mentioned properties of W. If

$$\int_0^a g(x)w(x)\,dx = \int_0^A g\circ W^{-1}(x)\,dx$$

is finite, then the primary Ψ of the composition $g\circ W^{-1} : [0,A] \mapsto R$ is a well-defined finite absolutely continuous function. Accordingly, there exist a unique smallest concave majorant $\bar{\Psi}$ of Ψ and a nonincreasing version $\bar{\psi}$ of the derivative of $\bar{\Psi}$. Put finally

$$P^{\searrow}g(x) = \bar{\psi}\circ W(x), \tag{4}$$

which is a well-defined nonincreasing function on I and

$$P_+^{\searrow}g = [(P^{\searrow}g)_+]_{|I}. \tag{5}$$

Below we construct the operators (4) and (5) for $\tilde{f}_\varphi = \tilde{f} + \varphi/w$ with various φ being elements of specified finite dimensional subspaces.

PROPOSITION **1** *If for a given X, $\tilde{f}(X,\cdot) \in \mathcal{H}$ and the integrals $\int_I w(x)\,dx$ and $\int_I dx/w(x)$ are finite, then*

$$Q\tilde{f} = P_+^{\searrow}\tilde{f}_c,$$

where c is a unique constant determined by

$$\int_I P_+^{\searrow}\tilde{f}_c = 1, \tag{6}$$

is the projection of $\tilde{f}(X,\cdot)$ onto (2).

It is known that $P^{\searrow}\tilde{f}$ describes the projection of $\tilde{f}_{|I}$ onto the convex cone $\mathcal{H}^{\searrow}_{|I}$ of nonincreasing functions in $\mathcal{H}_{|I} = L_2(I, w(x)\,dx)$. To the best of the author's knowledge, the result was originally established by Moriguti (1953) in the case of uniform weighting. The discrete version of $P^{\searrow}$, which can be effectively determined by the pool adjacent violators algorithm proposed in Ayer *et al.* (1955), was extensively applied in various problems of order restricted inference (see Robertson *et al.* (1988)). Furthermore, $\tilde{f}_{c|I} = (\tilde{f} + c/w)_{|I}$ with $c = (1 - \int_I \tilde{f})/\int_I (1/w)$ provides the projection of $\tilde{f}_{|I}$ onto the family of functions integrating to 1 (see Gajek (1986, Lemma 2)). Finally, it can be trivially concluded that $\tilde{f}_{|I}$ and $\tilde{f}_{+}$ are the projections of $\tilde{f}$ onto the families of functions supported on I and nonnegative ones, respectively. A nontrivial matter is composing the four operations so that we obtain the projections onto the set of functions (2) satisfying the respective conditions altogether. Observe that generally it is impossible to determine c in (6) analytically, because admitting strongly fluctuating weight functions may result in significant changes of c. Fortunately, the function

$$L(c) = \int_I P_+^{\searrow}\tilde{f}_c \tag{7}$$

is continuous and strictly increasing for $c \geq \sup\{c :\ L(c) = 0\}$. Therefore a numerical solution of (6) can be determined easily.

Note that a modification of Proposition 1 for the case of nondecreasing densities can be immediately derived once we replace the smallest concave majorant by the greatest convex minorant in the construction of $P^{\searrow}$. More generally, we can similarly treat the problem of estimating the density which is piecewise nondecreasing and nonincreasing, and the locations of respective peaks and valleys are known. However, one cannot usually find a practical justification for claiming that the density of interest has a number of precisely determined intervals of monotonicity. More practical applications can be found when considering the class of symmetric unimodal densities with a given, say 0, mode. Define $J = [-a, a]$ for some $a \in R_+$, and

$$\begin{aligned}
\mathcal{G} &= \{h \in \mathcal{H} :\ h \geq 0,\ \int h = 1,\ h = 0 \text{ on } J^c,\ h \text{ is sym-} \\
&\qquad\qquad \text{metric about 0 and nonincreasing on } R_+\}, \qquad (8)\\
w^s(x) &= w(x) + w(-x), \\
g^s_\varphi(x) &= \frac{g(x)w(x) + g(-x)w(-x) + \varphi(x) + \varphi(-x)}{w^s(x)},\quad g, \varphi \in \mathcal{H}. \qquad (9)
\end{aligned}$$

Proposition **2** *Given $g \in \mathcal{H}$, let $S_+^{\searrow}g$ be the function which is symmetric about 0, and defined as $P_+^{\searrow}g$ on R_+ for the symmetrized weight w^s instead of w. Under the hypotheses of Proposition 1, with I replaced by J, there is* *a unique number c satisfying*

$$\int_I S_+^{\searrow} \tilde{f}_c^s(x)\,dx = 1/2,$$

such that $S_+^{\searrow} \tilde{f}_c^s$ is the projection of $\tilde{f}$ onto (8).

More precise approximations of the parent density can be obtained if we impose expectation and/or some other moment constraints. The bias-reducing kernel estimates have some moments identical with their sample counterparts. The improvement procedures presented in Propositions 1 and 2 violate these nice properties, but we can recover them adding extra moment conditions for distance-minimizing problems. One should be aware though that the respective projections ensure the reduction of $WISE$ for densities whose moments coincide with those for the empirical distribution for the given X.

Generally, we additionally assume that f is a member of the family

$$\mathcal{F}^m = \{h \in \mathcal{F} : \int \phi_i(x)h(x)\,dx = m_i,\ i = 1, \ldots, k\}, \tag{10}$$

where ϕ_i are nonconstant, linearly independent, and such that all $\int_I \phi_i^2/w$, $i = 1, \ldots, k$, are finite. Note that for some choices of $m = (m_1, \ldots, m_k)$ the system of equations in (10) may have no solution. It can be concluded from Kemperman (1971) (see also Anastassiou (1993, Section 2.2)) that a sufficient condition for $\mathcal{F}^m$ being nonempty is that the moment vector m is an inner point of the convex hull of the image of the vector function $\Phi = (\Phi_1, \ldots, \Phi_k) : I \mapsto R^k$, with $\Phi_i(x) = \frac{1}{x}\int_0^x \phi_i(t)\,dt$, $i = 1, \ldots, k$, i. e.,

$$m \in \text{int conv}\{\Phi(x) :\ x \in I\}. \tag{11}$$

Note that if some ϕ_i are monotone, then the moment vector m of the uniform density supported on I is a border point of the convex hull. However, other border points generally correspond to mixtures of nonincreasing densities on I and a pole at the left end-point.

Proposition **3** *If $\int_I \phi_i^2/w < \infty$, $i = 1, \ldots, k$, and* (11) *holds, then there exists a sequence $\varphi_n \in \text{span}\{\phi_i :\ i = 0, \ldots, k\}$ with $\phi_0 \equiv 1$, $n = 1, 2, \ldots$, such that the unique projection $Q^m \tilde{f}$ of $\tilde{f}$ onto* (10) *can be represented as*

$$Q^m \tilde{f} = \lim_{n\to\infty} P_+^{\searrow} \tilde{f}_{\varphi_n}.$$

The proof is constructive and we write down explicit formulas for the coefficients $c_{i,n}$, $i = 0, \ldots, k$, in the representation of each $\varphi_n = \sum_{i=0}^k c_{i,n}\phi_i$, $n = 1, 2, \ldots$, in Section 4. One can also try to find $\varphi = \sum_{i=0}^k c_i\phi_i$ such that $P_+^{\searrow} \tilde{f}_{\varphi}$ satisfies the moment constraints and check if this is actually the desired projection (e. g., verifying a condition analogous to (17) below). However, it may generally happen that the moment conditions on $P_+^{\searrow}\tilde{f}_\varphi$ do not uniquely determine φ. Another possibility is that various systems of coefficients could lead to the same final form of projection. Analogous reservations concern the symmetric case. With the notation

$$\begin{aligned} \mathcal{G}^m &= \{h \in \mathcal{G} : \int \phi_i(x)h(x)\,dx = m_i,\ i = 1,\dots,k\}, \qquad (12)\\ \Phi_i^s(x) &= \frac{1}{x}\int_0^x [\phi_i(t) + \phi_i(-t)]\,dt, \qquad i = 1,\dots,k,\ x \in I, \end{aligned}$$

we have the following

Proposition 4 *If $\int_J \phi_i^2/w < \infty$, $i = 1,\dots,k$, and $m \in \operatorname{int}\operatorname{conv}\{\Phi^s(x) : x \in I\}$, then there is a sequence $\varphi_n \in \operatorname{span}\{\phi_i : i = 0,\dots,k\}$, $n = 1,2,\dots$, such that the projection $S^m\tilde{f}$ of $\tilde{f}$ onto* (12) *is the symmetric function defined on R_+ by*

$$S^m\tilde{f} = \lim_{n\to\infty} S_+^{\searrow}\tilde{f}^s_{\varphi_n}.$$

3 PROOFS

The idea of proofs consists in decomposing the convex sets (2), (8), (10) and (12) into the intersections of simpler ones for which the respective projections are more tractable. The latter have the following forms:

$$\begin{aligned} \mathcal{F}^\mu &= \{h \in \mathcal{H} : \int \phi(x)h(x)\,dx = \mu,\ h = 0 \text{ on } I^c\}, &(13)\\ \mathcal{F}_+^{\searrow} &= \{h \in \mathcal{H} :\ h \geq 0 \text{ and is nonincreasing on } I,\ h = 0 \text{ on } I^c\}, &(14)\\ \mathcal{G}_+^{\searrow} &= \{h \in \mathcal{H} :\ h \geq 0,\ h = 0 \text{ on } J^c,\ h \text{ is symmetric} &(15)\\ &\qquad \text{about 0 and nonincreasing on } R_+\}. \end{aligned}$$

The respective projections are presented in Lemmas 1, 2 and 3. Then we describe the Boyle-Dykstra algorythm whose convergence is stated in Lemma 4. This is applied for establishing the final forms of projections considered in Section 2. We provide a detailed proof of Proposition 1 and outline the proofs of the other results.

Lemma 1 *If $\phi/w \in \mathcal{H}_{|I}$ then for any $g \in \mathcal{H}$*

$$P^\mu g = (g + b\frac{\phi}{w})_{|I}, \qquad (16)$$

with

$$b = (\mu - \int_I \phi g)/\int_I \frac{\phi^2}{w},$$

is the projection of g onto (13).

Proof. This is a slight modification of Lenic (1994) and we write it down for completeness. We first observe that b is well defined since

$$|\int_I \phi g| \leq \left(\int_I \frac{\phi^2}{w} \int_I g^2 w\right)^{1/2} < \infty,$$

and therefore $P^\mu g$ is actually an element of $\mathcal{F}^\mu$. We apply a condition characterizing the projections onto closed convex subsets of Hilbert spaces

$$\forall h \in \mathcal{F}^\mu \qquad (P^\mu g - g, P^\mu g - h) \leq 0, \tag{17}$$

(see, e. g., Balakrishnan (1976, p. 10)) which in our case is satisfied since

$$\int (P^\mu g - g)(P^\mu g - h)w = b\int_I (P^\mu g - h)\phi = 0. \quad \square$$

Here I can be replaced by any Borel set in R, and so by S in particular. Note that the correction in (16) is inversely proportional to the weight. This means that the greater significance is put on the accuracy of approximating g in some regions the less this is modified there.

Lemma **2** *The* $P_+^{\searrow} g$ *defined in* (5) *is the projection of* g *onto* (14).

Proof. We first note that (14) is a convex cone in $\mathcal{H}$. Therefore it suffices to verify two conditions

$$\forall h \in \mathcal{F}_+^{\searrow} \qquad (g,h) \leq (P_+^{\searrow} g, h), \tag{18}$$

$$(g, P_+^{\searrow} g) = (P_+^{\searrow} g, P_+^{\searrow} g), \tag{19}$$

uniquely characterizing projections onto convex cones (see, e. g., Balakrishnan (1976, p. 11–12)). We shall exploit the fact that $P^{\searrow} g$ is the projection of $g_{|I}$ onto the convex cone $\mathcal{H}_{|I}^{\searrow}$ of nonincreasing (not necessarily nonnegative) functions in $\mathcal{H}_{|I}$ which implies that

$$\forall h \in \mathcal{H}_{|I}^{\searrow} \qquad \int_I ghw \leq \int_I P^{\searrow} g\, hw, \tag{20}$$

$$\int_I gP^{\searrow} g\, w = \int_I (P^{\searrow} g)^2 w. \tag{21}$$

By (20), for all $h \in \mathcal{F}_+^{\searrow}$ we have

$$\int ghw \leq \int_I P^{\searrow} g\, hw \leq \int_I (P^{\searrow} g)_+ hw = \int P_+^{\searrow} g\, hw,$$

which gives (18). Set $\alpha = \sup\{x \in I :\ P^{\searrow} g(x) > 0\}$. If $P^{\searrow} g$ is nonpositive then (19) holds trivially. If $\alpha = a$ then (19) follows immediately from (21). So assume that α is an interior point of I. Referring to the construction of $P^{\searrow}g$ we observe that $W^{-1}(I) = [0, A]$ can be split into two sequences of open and closed subintervals, where $\bar{\Psi} > \Psi$ (and $\bar{\psi}$ is constant on each of the respective subintervals) and $\bar{\Psi} = \Psi$. Since α is a point at which $P^{\searrow}g = \bar{\psi} \circ W$ actually decreases, we have $\bar{\psi}(W(\alpha)) = \psi(W(\alpha))$. In particular, this implies that $P^{\searrow}g_{|[0,\alpha]}$ is the $\mathcal{H}_{|[0,\alpha]}$-projection of $g_{|[0,\alpha]}$ onto the cone of nonincreasing functions. Applying the analogue of (21) with a replaced by α, we conclude that

$$\int g P_+^{\searrow} g\, w = \int_0^\alpha g P^{\searrow} g\, w = \int_0^\alpha (P^{\searrow} g)^2 w = \int (P_+^{\searrow} g)^2 w$$

which provides (21) and completes the proof. □

Lemma **3** *For $S_+^{\searrow}$ defined in Proposition 2, and*

$$g_0^s(x) = \frac{g(x)w(x) + g(-x)w(-x)}{w^s(x)}, \quad x \in I,\ g \in \mathcal{H}.$$

(cf. equation (9)*), the projection of g onto* (15) *is defined as $S_+^{\searrow} g_0^s$.*

Proof. Since $\mathcal{G}_+^{\searrow}$ is a convex cone, we check two criteria analogous to (18)–(19) from the proof of Lemma 2. Also, some conclusions of Lemma 2 will be exploited here. The former criterion is fulfilled, because for arbitrary $h \in \mathcal{G}_+^{\searrow}$ yields

$$\begin{aligned} (g, h) &= \int_J g\, h\, w = \int_I g_0^s\, h\, w^s \\ &\le \int_I S_+^{\searrow} g_0^s\, h\, w^s = \int_J S_+^{\searrow} g_0^s\, h\, w = (S_+^{\searrow} g_0^s, h). \end{aligned}$$

The above inequality is a consequence of the fact that $S_+^{\searrow}$ is the projection operator onto the cone of nonnegative nonincreasing functions on I with respect to weight w^s. By the same arguments, we have

$$\begin{aligned} (g, S_+^{\searrow} g_0^s) &= \int_J g\, S_+^{\searrow} g_0^s\, w = \int_I g_0^s\, S_+^{\searrow} g_0^s\, w^s \\ &= \int_I (S_+^{\searrow} g_0^s)^2 w^s = \int_J (S_+^{\searrow} g_0^s)^2 w = (S_+^{\searrow} g_0^s, S_+^{\searrow} g_0^s), \end{aligned}$$

which proves the lemma. □

We now present the Boyle-Dykstra algorithm that allows us to combine the projections P_i onto convex closed sets $\mathcal{H}_i$, $i = 1, \ldots, k$, so as to derive sequences converging to the projection P onto the nonempty intersection of the $\mathcal{H}_i$. The algorithm consists of infinite sequence of steps, and each step consists of k substeps. Suppose that h is to be projected and put $h_{1,0} = h$, and $\Delta_{0,i} = 0$, $i = 1, \dots, k$. At the step n, $n = 1, 2, \dots$, and for each substep $i = 1, \dots, k$, we consecutively determine the projection

$$h_{n,i} = P_i(h_{n,i-1} + \Delta_{n-1,i})$$

and the respective correction

$$\Delta_{n,i} = (h_{n,i-1} + \Delta_{n-1,i}) - P_i(h_{n,i-1} + \Delta_{n-1,i}) = h_{n,i-1} + \Delta_{n-1,i} - h_{n,i},$$

$i = 1, \dots, k$, and write finally $h_{n+1,0} = h_{n,k}$. The latter, together with all $\Delta_{n,i}$, $i = 1, \dots, k$, are applied in step $n+1$.

LEMMA **4** (Boyle and Dykstra 1986). *Each of the sequences $h_{n,i}$, $i = 1, \dots, k$, tends in the norm to Ph, as $n \to \infty$. Moreover, if any of $\mathcal{H}_i$ is either a linear subspace or a translate of a linear subspace, then we can put $\Delta_{n,i} = 0$, $n = 1, 2, \dots$, and the conclusion still holds.*

Proof of Proposition 1. Observe that we can write $\mathcal{F} = \mathcal{F}_1 \cap \mathcal{F}_2$ with

$$\mathcal{F}_1 = \{h \in \mathcal{H} : \int h = 1,\ h = 0 \text{ on } I^c\}$$

and $\mathcal{F}_2 = \mathcal{F}_+^{\searrow}$ (see formulas (13), (14)). For the above decomposition we apply the Boyle-Dykstra algorithm to show that the outcome of each step has the form $P_+^{\searrow}\tilde{f}_{c_n}$ for some constant c_n. Notice first that $\mathcal{F}_1$ is a translate of a linear subspace in $\mathcal{H}$ and, by Lemma 4, we can put $\Delta_{1,n} = 0$ for $n = 1, 2, \dots$. From Lemmas 1 and 2 we conclude that

$$\begin{aligned}
\tilde{f}_{1,1} &= P^1\tilde{f} = (\tilde{f} + b_1/w)_{|I}, \\
\tilde{f}_{1,2} &= P_+^{\searrow}\tilde{f}_{1,1} = P_+^{\searrow}(\tilde{f} + b_1/w) = P_+^{\searrow}\tilde{f}_{b_1}, &(22)\\
\Delta_{1,2} &= (\tilde{f} + b_1/w)_{|I} - P_+^{\searrow}(\tilde{f} + b_1/w). &(23)
\end{aligned}$$

Put $c_1 = b_1$ and suppose that for some $n \geq 1$ functions $\tilde{f}_{n,2}$ and $\Delta_{n,2}$ have the forms (22) and (23), respectively, with b_1 replaced by some c_n. Proceeding to step $n+1$, we obtain

$$\begin{aligned}
\tilde{f}_{n+1,1} &= P_+^{\searrow}(\tilde{f} + c_n/w) + (b_{n+1}/w)_{|I}, &(24)\\
\tilde{f}_{n+1,2} &= P_+^{\searrow}(\tilde{f}_{n+1,1} + \Delta_{n,2}) = P_+^{\searrow}(\tilde{f} + c_{n+1}/w) = P_+^{\searrow}\tilde{f}_{c_{n+1}}, &(25)
\end{aligned}$$

with $c_{n+1} = c_n + b_{n+1}$. By Lemma 4, both (24) and (25) tend to $Q\tilde{f}$, as $n \to \infty$. We analyze more closely the latter sequence whose elements are determined up to single real parameters c_n.

We claim that $Q\tilde{f} = P_+^{\searrow}\tilde{f}_c$ with a uniquely determined parameter $c = \lim_{n\to\infty} c_n$. We first recall the representation $P^{\searrow}\tilde{f}_c = \bar{\psi}_c \circ W$ with

$$\begin{aligned}
\bar{\psi}_c(x) &= \inf_{x\le v\le A} \sup_{0\le u\le x} \frac{1}{v-u}\int_u^v \tilde{f}_c \circ W^{-1} && (26)\\
&= \inf_{x\le v\le A} \sup_{0\le u\le x} \frac{1}{v-u}\left(\int_u^v \tilde{f}\circ W^{-1} + c\int_u^v \frac{1}{w\circ W^{-1}}\right), \quad 0\le x\le A,
\end{aligned}$$

(cf. equation (4) and, e. g., Mammen (1991a, p. 726), respectively). Since for arbitrary u, v and $c \to +\infty$ the expression in parentheses tends to $+\infty$, we conclude that $P^{\searrow}\tilde{f}_c \to +\infty$ pointwise and so does $P_+^{\searrow}\tilde{f}_c$. Letting $c \to -\infty$, we see that $P^{\searrow}\tilde{f}_c \to -\infty$ and $P_+^{\searrow}\tilde{f}_c \to 0$. This implies boundedness of c_n.

Applying (26) again we can deduce that $P_+^{\searrow}\tilde{f}_{c'} \le P_+^{\searrow}\tilde{f}_{c''}$ if $c' \le c''$. In fact, a stronger assertion holds: if $c' < c''$ and $P_+^{\searrow}\tilde{f}_{c''} \not\equiv 0$, then $P_+^{\searrow}\tilde{f}_{c'} < P_+^{\searrow}\tilde{f}_{c''}$ on a set of positive measure. On the contrary, suppose that $P_+^{\searrow}\tilde{f}_{c'} = P_+^{\searrow}\tilde{f}_{c''}$ which is equivalent to saying that the smallest concave majorants $\bar{\Psi}_{c'}$ and $\bar{\Psi}_{c''}$, appearing in the constructions of these functions, coincide at least up to a common positive maximum point $W(\beta) \in (0, A]$, say. However, arguing as in the proof of Lemma 2, we get

$$\begin{aligned}
\bar{\Psi}_{c''}(W(\beta)) - \bar{\Psi}_{c'}(W(\beta)) &= \Psi_{c''}(W(\beta)) - \Psi_{c'}(W(\beta))\\
&= (c'' - c')\beta > 0,
\end{aligned}$$

which contradicts our claim. Thus $P_+^{\searrow}\tilde{f}_{c'}$ and $P_+^{\searrow}\tilde{f}_{c''}$ are actually different and $L(c') < L(c'')$ (see equation (7)). This enables us to conclude that c_n, $n = 1, 2, \ldots$, determined by the algorythm have a unique concentration point, and there is at most one c such that $L(c) = 1$.

It it now sufficient to show that $P_+^{\searrow}\tilde{f}_{c_n} \to P_+^{\searrow}\tilde{f}_c$, as $c_n \to c$. To this end, we refer again to some steps of construction of $P_+^{\searrow}$. If $c_n \ge c$ then

$$\begin{aligned}
\Psi_{c_n}(x) &= \int_0^x \tilde{f}_{c_n} \circ W^{-1}(t)\,dt && (27)\\
&\le \int_0^x \tilde{f}_c \circ W^{-1}(t)\,dt + (c_n - c)\int_0^A \frac{dt}{w\circ W^{-1}(t)}\\
&= \Psi_c(x) + (c_n - c)a
\end{aligned}$$

and, in consequence, for the respective smallest concave majorants we have

$$\bar{\Psi}_c \le \bar{\Psi}_{c_n} \le \bar{\Psi}_c + (c_n - c)a \to \bar{\Psi}_c \qquad (28)$$

uniformly on $[0, A]$, as $n \to \infty$. If $c_n \le c$ the reversed inequalities in (27) and (28) and the same final conclusion hold. The limiting $\bar{\Psi}_c$ uniquely determines $P_+^{\searrow}\tilde{f}_c$ that integrates to 1. $\square$

Proof of Proposition 2 is a modification of the preceding one. We can write

$$\mathcal{G} = \mathcal{G}_1 \cap \mathcal{G}_2 = \{h \in \mathcal{H} : \int h = 1,\ h = 0 \text{ on } J^c\} \cap \mathcal{G}_+^{\searrow}.$$

In the first step of the Boyle-Dykstra procedure we obtain

$$\tilde{f}_{1,1} = (\tilde{f} + b_1/w)_{|J}$$

with $b_1 = \left(1 - \int_J \tilde{f}\right) / \int_J \frac{1}{w}$ (see Lemma 1), and

$$\begin{aligned} \tilde{f}_{1,2} &= S_+^{\searrow}(\tilde{f} + b_1/w)_0^s = S_+^{\searrow}\tilde{f}_{b_1}^s, \\ \Delta_{1,2} &= (\tilde{f} + b_1/w)_{|J} - S_+^{\searrow}\tilde{f}_{b_1}^s \end{aligned}$$

(see Lemma 3). The second step provides

$$\tilde{f}_{2,1} = S_+^{\searrow}\tilde{f}_{b_1}^s + (b_2/w)_{|J},$$

for some b_2 and

$$\tilde{f}_{2,2} = S_+^{\searrow}(\tilde{f}_{2,1} + \Delta_{1,2}) = S_+^{\searrow}\left(\tilde{f} + (b_1 + b_2)/w\right)_0^s = S_+^{\searrow}\tilde{f}_{b_1+b_2}^s.$$

One can easily check that $\tilde{f}_{n,2} = S_+^{\searrow}\tilde{f}_{c_n}^s$, $n \geq 1$, for $c_n = \sum_{i=1}^n b_i$. Note that

$$\begin{aligned} \tilde{f}_{c_n}^s(x) &= \frac{\tilde{f}(x)w(x) + \tilde{f}(-x)w(-x)}{w(x) + w(-x)} + \frac{2c_n}{w(x) + w(-x)} \\ &= \tilde{f}_0^s(x) + \frac{2c_n}{w^s(x)}, \qquad x \in I, \end{aligned}$$

and $S_+^{\searrow}$ projects functions supported on I onto (14) with respect to weight function w^s. If we substitute $S_+^{\searrow}$, $\tilde{f}_0^s$, $2c_n$ and w^s for $P_+^{\searrow}$, $\tilde{f}$, c_n and w, respectively, in the second part of the proof of Proposition 1 we immediately establish the desired conclusion. □

Proof of Proposition 3. It suffices to show that the elements of one of the sequences determined by the Boyle-Dykstra algorithm can be written in the form $P_+^{\searrow}\left(\tilde{f} + \frac{1}{w}\sum_{i=1}^{k+1} c_{i,n}\phi_{i-1}\right)$, $n = 1, 2, \ldots$, for a properly chosen representation of (10) as the intersection of convex sets with a simpler structure. We propose setting $\mathcal{F}^m = \bigcap_{i=1}^{k+2} \mathcal{F}_i$, where

$$\mathcal{F}_i = \mathcal{F}^{m_{i-1}} = \{h \in \mathcal{H} : \int h\,\phi_{i-1} = m_{i-1},\ h = 0 \text{ on } I^c\}, \quad i = 1, \ldots, k+1,$$

with $\phi_0 \equiv 1 = m_0$ and $\mathcal{F}_{k+2} = \mathcal{F}_+^{\searrow}$ (see formulas (13), (14)). By Lemmas 1 and 2,

$$\begin{aligned}
\tilde{f}_{1,l} &= \left(\tilde{f} + \frac{1}{w}\sum_{i=1}^{l} c_{1,i}\psi_i\right)_{|I}, \qquad l = 1,\ldots,k+1, && (29)\\
\tilde{f}_{1,k+2} &= P_+^{\searrow}\left(\tilde{f} + \frac{1}{w}\sum_{i=1}^{k+1} c_{1,i}\psi_i\right), && (30)\\
\Delta_{1,k+2} &= \left(\tilde{f} + \frac{1}{w}\sum_{i=1}^{k+1} c_{1,i}\psi_i\right) - P_+^{\searrow}\left(\tilde{f} + \frac{1}{w}\sum_{i=1}^{k+1} c_{1,i}\psi_i\right). && (31)
\end{aligned}$$

Note that we can put $\Delta_{n,i} = 0$ for $i = 1,\ldots,k+1$ and all $n \geq 1$. Replacing 1 by an arbitrarily given natural number n in the indices of the left-hand sides of (29), (30), (31) as well as in those of the coefficients of the combinations of the respective right-hand sides, we determine

$$\begin{aligned}
\tilde{f}_{n+1,l} &= P_+^{\searrow}\left(\tilde{f} + \frac{1}{w}\sum_{i=1}^{k+1} c_{n,i}\psi_i\right) + \left(\frac{1}{w}\sum_{i=1}^{l} b_{n+1,i}\psi_i\right)_{|I}, \; l = 1,\ldots,k+1,\\
\tilde{f}_{n+1,k+2} &= P_+^{\searrow}\left(\tilde{f}_{n+1,k+1} + \Delta_{n,k+2}\right) = P_+^{\searrow}\left(\tilde{f} + \frac{1}{w}\sum_{i=1}^{k+1} c_{n+1,i}\psi_i\right),
\end{aligned}$$

with $c_{n+1,i} = c_{n,i} + b_{n+1,i}$, $i = 1,\ldots,k+1$. The last sequence meets our requirements. □

The same proof works for Proposition 4 when we substitute $\mathcal{G}_+^{\searrow}$ for $\mathcal{F}_+^{\searrow}$. The details are left to the reader. Versions of the above algorithms which are suitable for numerical implementations will be presented in Section 4.

4 ALGORITHMS AND REMARKS

An algorithm for projecting a density estimate $\tilde{f}$ onto the nondecreasing densities with given support I and moment conditions $\int_I \phi_i f = m_i$, $i = 0,\ldots,k$. (Here I is an arbitrary finite interval, $\phi_0 \equiv 1$ and $m_0 = 1$.)

Determine first

$$\begin{aligned}
A_{j,i} &= \int_I \phi_j\phi_i/w, && 0 \leq j \leq i \leq k,\\
\chi_i(x) &= \phi_i(x)/w(x), && x \in I.
\end{aligned}$$

Put

$$n = 0,$$

$$\begin{aligned} c_{0,i} &= 0, && i = 0,\ldots,k, \\ \delta_{0,i} &= m_i - \int_I \tilde{f}\phi_i, && i = 0,\ldots,k. \end{aligned}$$

Then

$$\begin{aligned} n &= n+1, && && (32) \\ b_{n,i} &= \left(\delta_{n-1,i} - \sum_{j=0}^{i-1} b_{n,j}A_{j,i}\right) / A_{i,i}, && i = 0,\ldots,k, \\ c_{n,i} &= c_{n-1,i} + b_{n,i} && i = 0,\ldots,k, \\ \tilde{f}_n(x) &= P_+^{\searrow}\left(\tilde{f} + \sum_{i=0}^{k} c_{n,i}\chi_i\right)(x), && x \in I, && (33) \\ \delta_{n,i} &= m_i - \int_I \tilde{f}_n\phi_i, && i = 0,\ldots,k. \\ \textbf{If} \quad & \delta_{n,i} = 0, && i = 0,\ldots,k, && (34) \\ \textbf{then} \quad & Q^m\tilde{f} = \tilde{f}_{n|I}, \quad \textbf{stop} \\ \textbf{else} \quad & \textbf{go to } (\mathbf{32}). \end{aligned}$$

An algorithm for projecting a density estimate $\tilde{f}$ onto the symmetric unimodal densities with given center of symmetry c and support $J = [c-a, c+a]$ and moment conditions $\int_J \phi_i f = m_i$, $i = 0,\ldots,k$. (Here $\int_J \phi_0 f = \int_J f = m_0 = 1$.) Determine first

$$\begin{aligned} A_{j,i} &= \int_J \phi_j\phi_i/w, && 0 \le j \le i \le k, \\ w^s(x) &= w(c+x) + w(c-x), && x \in [0,a], && (35) \\ \tilde{f}^s(x) &= [\tilde{f}(c+x)w(c+x) + \tilde{f}(c-x)w(c-x)]/w^s(x), && x \in [0,a], \\ \phi_i^s(x) &= \phi_i(c+x) + \phi_i(c-x), && x \in [0,a],\ i = 0,\ldots,k, \\ \chi_i^s(x) &= \phi_i^s(x)/w^s(x), && x \in [0,a],\ i = 0,\ldots,k. \end{aligned}$$

Put

$$\begin{aligned} n &= 0, \\ c_{0,i} &= 0, && i = 0,\ldots,k, \\ \delta_{0,i} &= m_i - \int_J \tilde{f}\phi_i, && i = 0,\ldots,k. \end{aligned}$$

Then

$$n = n+1, \qquad (36)$$

$$
\begin{array}{rcll}
b_{n,i} & = & \left(\delta_{n-1,i} - \sum_{j=0}^{i-1} b_{n,j} A_{j,i}\right) / A_{i,i}, & i = 0, \ldots, k, \\
c_{n,i} & = & c_{n-1,i} + b_{n,i} & i = 0, \ldots, k, \\
\tilde{f}_n(x) & = & S_+^{\searrow}\left(\tilde{f}^s + \sum_{i=0}^{k} c_{n,i}\chi_i^s\right)(x), & x \in [0, a], \qquad (37) \\
\delta_{n,i} & = & m_i - \int_0^a \tilde{f}_n \phi_i^s, & i = 0, \ldots, k. \\
\textbf{If} & & \delta_{n,i} = 0, & i = 0, \ldots, k, \qquad (38) \\
\textbf{then} & & S^m \tilde{f}(x) = \left\{ \begin{array}{ll} \tilde{f}_n(x-c), & c \le x \le c+a, \\ \tilde{f}_n(c-x), & c-a \le x \le c, \\ 0, & x \in J^c, \end{array} \right. & \textbf{stop} \\
\textbf{else} & & \textbf{go to } (\mathbf{36}). &
\end{array}
$$

Clearly, conditions (34) and (38) imply that $\tilde{f}_n$ would satisfy all moment and shape conditions for some finite n, something which cannot be expected in general. One could make the procedures finite either by adding a condition on the maximal number of steps or by relaxing (34) and (38). For instance, the algorithms could stop when the distance in a chosen metric of $(\delta_{n,0}, \ldots, \delta_{n,k})$ to the origin in R^{k+1} (or to $(\delta_{n-1,0}, \ldots, \delta_{n-1,k})$) were less than a given ϵ.

The only nonelementary operations of the algorithms, except of the numerical integration, are (33) and (37). These can be well approximated by means of the weighted pool adjacent violators algorithm. To get the former one, we determine the isotonic regression of $-\left(\tilde{f} + \sum_{i=0}^{k} c_{n,i}\chi_i\right)$ with respect to weight w over a sufficiently dense grid in I, then change the sign and reject the negative values. For the latter, we apply the symmetrized weight (35).

However, in the case of symmetric weighting, which is the most natural here, $w^s = w$ and the argument of the right-hand side of (37) reduces to

$$
\tilde{f}_c(x) = \frac{\tilde{f}(x) + \tilde{f}(-x)}{2} + \frac{1}{w(x)} \sum_{i=0}^{k} c_{n,i} \frac{\phi_i(x) + \phi_i(-x)}{2}.
$$

Then determining $S^m \tilde{f}$ simply consists in folding $\tilde{f}$ and ϕ_i, $i = 0, \ldots, k$, about 0, applying the former algorithm, and unfolding the solution.

The most natural moment condition are $\mathrm{E}(X^i) = m_i$, $i = 0, \ldots, k$. It is well known that $\int_0^1 x^i p(x)\,dx = \mu_i$, $i = 0, \ldots, l$, for some nonnegative, not necessarily probabilistic density function p on $[0, 1]$ if the following matrices are positive definite

$$
[\mu_{i+j-2}]_{i,j=1,\ldots,\lfloor l/2+1 \rfloor}, \qquad (39)
$$

$$
[\mu_{i+j-1}]_{i,j=1,\ldots,\lfloor (l+1)/2 \rfloor}, \qquad (40)
$$

$$[\mu_{i+j-2} - \mu_{i+j-1}]_{i,j=1,\ldots,\lfloor (l+1)/2 \rfloor}, \tag{41}$$

$$[\mu_{i+j-1} - \mu_{i+j}]_{i,j=1,\ldots,\lfloor l/2 \rfloor}, \tag{42}$$

where $\lfloor \cdot \rfloor$ denotes the floor of a number (see, e. g., Karlin and Studden 1966, Chapter 4). If we consider nonincreasing densities supported on the interval $I = [c, c+a]$ with given first k standard moments m_i, $i = 1, \ldots, k$, then, applying (11) and linear transformations, we conclude that the moment conditions are not contradictory if (39)–(42) are positive definite with $l = k$ and

$$\mu_i = \frac{i+1}{a^i} \sum_{j=0}^{i} \binom{i}{j} (-c)^{i-j} m_j, \quad i = 0, \ldots, k. \tag{43}$$

The possible standard moments $m_1, \ldots, m_k$ of symmetric and unimodal $[c - a, c + a]$-supported densities satisfy

$$\sum_{j=0}^{2i+1} \binom{2i+1}{j} (-1)^{j-1} c^{2i+1-j} m_j = 0, \quad 1 \le 2i+1 \le k,$$

together with positive definiteness of (39)–(42) for $l = [k/2]$ and

$$\mu_i = \frac{2i+1}{2a^{2i}} \sum_{j=0}^{2i} \binom{2i}{j} (-1)^j c^{2i-j} m_j, \quad 1 \le 2i \le k. \tag{44}$$

Obviously, the above algorithms with $k = 0$ provide the projections onto respective classes of densities without moment constraints. Moreover, if we assume the uniform weighting and if the original estimate $\tilde{f}$ is nonnegative and $\int \tilde{f} \le 1$, then both the algorithms terminate in one step. Then $\tilde{f}_{1,1} \ge \tilde{f}$ is a proper density function and so is $\tilde{f}_{1,2}$. This is because $P^{\searrow}\tilde{f}$ preserves the value of $\int \tilde{f} w$ and the sign of the one-signed function $\tilde{f}$, since $P^{\searrow}\tilde{f}$ is defined by averaging $\tilde{f}$ on some intervals.

We can slightly change the algorithms if the conditions of monotonicity and unimodality are not imposed. Then we only need to replace the operators $P_+^{\searrow}$ and $S_+^{\searrow}$ in (33) and (37), respectively, by taking the positive parts of the respective arguments. In these cases, we also modify the conditions for standard moments of densities by dropping $i + 1$ and $2i + 1$ from the numerators of the factors standing in front of the sums in (43) and (44), respectively.

ACKNOWLEDGEMENTS

The research was done under the support of Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonometrischer Prozesse" when the author was visiting Insititute of Statistics and Econometry, Humboldt University in Berlin. This was also partially supported by the State Committee for Scientific

Research (KBN) Grant 2 P03A 014 13. The author is grateful to the Editor, an Associate Editor, a referee, and Philippe Barbe for their valuable comments that allowed him to improve the presentation of the paper.

REFERENCES

Anastassiou, G. A. (1993). *Moments in Probability and Approximation Theory.* Longman Sci. Tech., Harlow, UK.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* 26, 641–647.

Balakrishnan, A. V. (1976). *Applied Functional Analysis.* Springer-Verlag, New York.

Barbe, P. (1998). Statistical analysis of mixtures and the empirical probability measure. *Acta Appl. Math.* 50, 253–340.

Barlow, R. E., and van Zwet, W. R. (1970). Asymptotic properties of isotonic estimators for the generalized failure rate function. Part I: strong consistency. *Nonparametric Techniques in Statistical Inference* (M.L. Puri, *ed.*), 159–173, Cambridge Univ. Press.

Barlow, R. E., and van Zwet, W. R. (1971). Comparisons of several nonparametric estimators of the failure ratio function. *Operations Research and Reliability. Proc. of the NATO Conf. Turin, 1969*, 375–399, Gordon Breach, New York.

Basu, A., Harris, I. R., and Basu, S. (1997). Minimum distance estimation: the approach using density based distances. *Robust Inference. Handbook of Statistics* Vol. 15 (G. S. Maddala, and C. R. Rao, *eds.*), 21–48, North-Holland, Amsterdam.

Beran, R. J. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* 5, 445–463.

Birgé, L. (1987*a*). Estimating a density under order restrictions: nonasymptotic minimax risk. *Ann.Statist.* 15, 995–1012.

Birgé, L. (1987*b*). On the risk of histograms for estimating decreasing densities. *Ann. Statist.* 15, 1013–1022.

Birgé, L. (1989). The Grenander estimator: a nonasymptotic approach. *Ann. Statist.* 17, 1532–1549.

Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* 25, 970–981.

Boyle, J. P., and Dykstra, R. L. (1986). A method for finding projections onto the intersection of convex sets in Hilbert spaces. *Advances in Order Restricted Statistical Inference* (R. L. Dykstra, T. Robertson., and F. T. Wright, *eds.*), 28–47, Springer-Verlag, New York.

Deely, J. J., and Kruse, R. L. (1968). Construction of sequences estimating the mixing distribution. *Ann. Math. Statist.* 39, 286–288.

Delecroix, M., and Thomas-Agnan, C. (1998). Kernel and spline smoothing under shape restrictions. *Smoothing and Regression. Approaches, Computation and Application* (M. Schimek, *ed.*), Wiley, New York, to appear.

Devroye, L. (1987). *A Course in Density Estimation.* Birkhäuser, Boston.

Fougères, A.-L. (1997). Estimation de densités unimodales. *Canad. J. Statist.* 25, 375–387.

Friedman, J., and Tibshirani, R. J. (1984). The monotone smoothing of scatter plots.

Technometrics 26, 243–250.
Gajek, L. (1986). On improving density estimators which are not bona fide functions. *Ann. Statist.* 14, 1612–1618.
Gajek, L. (1987). Estimation of a density and its derivatives by the minimum distance method (in Polish). *Scient. Bull. Łódź Techn. Univ.* No. 533.
Grenander, U. (1956). On the theory of mortality measurement. Part II. *Skand. Akt.* 39, 125–153.
Groeneboom, P. (1985). Estimating a monotone density. *Proc. Berkeley Conf. in Honor of Neyman and Kiefer* Vol. 2, 529–555, Wadsworth, Inc.
Groeneboom, P., and Pyke, R. (1983). Asymptotic normality of statistics based on the convex minorants of empirical distribution functions. *Ann. Probab.* 11, 328–345.
Kałuszka, M., and Lenic, A. (1995). On improving density estimates which are not bona fide functions. *Disscuss. Math. Algebra Stochastic Methods* 15, 283–290.
Karlin, S., and Studden, W. J. (1966). *Tchebycheff Systems: with Applications in Analysis and Statistics.* Wiley, New York.
Kemperman, J. H. B. (1971). Moment problems with convexity conditions. *Optimiz--ing Methods in Statistics* (J. S. Rustagi, *ed.*), 115–178, Academic Press, New York.
Kiefer, J., and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. Verw. Gebiete* 34, 73–85.
Kintchine, A. Y. (1938). On unimodal distributions. *Izv. Nauchno-Issled. Inst. Mat. Mech. Tomsk. Gos. Univ* 2, 1–7.
Lenic, A. (1994). *Projection Methods in Nonparametric Density Estimation* (in Polish). Ph.D. thesis, Inst. of Math., Polish Acad. Sci.
Mammen, E. (1991*a*). Estimating a smooth monotone regression function. *Ann. Statist.* 19, 724–740.
Mammen, E. (1991*b*). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* 19, 741–759.
Mammen, E., and Thomas-Agnan, C. (1998). Smoothing splines and shape restrictions. *Scand. J. Statist.*, to appear.
Moriguti, S. (1953). A modification of Schwarz's inequality with applications to distributions. *Ann. Math. Statist.* 24, 107–113.
Mukerjee, H. (1988). Monotone nonparametric regression. *Ann. Statist.* 16, 741–750.
Prakasa Rao, B. L. S. (1969). Estimation of unimodal density. *Sankhyā A* 31, 23–36.
Robertson T., Wright F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, Wiley, New York.
Schuster, E. F. (1987). Identifying the closest symmetric distribution and density functions. *Ann. Statist.* 12, 867–874.
Stefafyuk, A. P. (1979). Convergence rate of a class of probability density estimates (in Russian). *Avtomat. i Telemekh.* 11, 187–192.
Utreras, F. I. (1985). Smoothing noisy data using monotonicity constraints. *Numerische Mathematik* 47, 611–625.
Vapnik, V. (1982). *Estimation of Dependencies Based on Empirical Data.* Springer-Verlag, New York.
Vapnik, V., and Stefanyuk, A. P. (1978). Nonparametric methods of probability density recovery (in Russian). *Avtomat. i Telemekh.* 8, 38–52.
Villalobos, M., and Wahba, G. (1987). Inequality constrained multivariate smooth-

ing splines with application to the estimation of posterior probabilities. *J. Amer. Statist. Assoc.* 82, 239–248.
Wright, F. T. (1982). Monotone regression estimates for grouped observations. *Ann. Statist.* 10, 278–286.

Institute of Mathematics
Polish Academy of Sciences
Chopina 12
87100 Toruń, Poland

email: trychlik@impan.gov.pl
phone: (48-52) 3425530
fax: (48-22) 6293997